HAIRPIN PEPTIDES WITH A NOVEL STRUCTURAL MOTIF AND METHODS RELATING THERETO

FIELD OF THE INVENTION

The present invention relates in general to protein chemistry, and more specifically to the identification and characterization of a novel small peptide motif with stable conformation, as well as to libraries of conformationally-constrained peptides and methods of generating and screening such libraries for biological and pharmaceutical uses.

BACKGROUND OF THE INVENTION

Structure-Activity Relationship (SAR) studies provide valuable insights for understanding intermolecular interactions between bioactive molecules. In their natural states, bioactive molecules often adopt unique, conformationally-constrained structures in order to recognize and bind to their binding partners, to form a molecular complex therewith, and in turn to elicit specific activities. In particular, protein-protein interactions are crucial events involved in most biological and pathological processes, and are therefore logical targets for drug design. Important protein-protein interactions occur between such binding partners as enzyme-substrate, ligand-receptor, and antigen-antibody complexes.

One of the revolutionary advances in drug discovery is the development of combinatorial libraries. Combinatorial libraries are collections of different molecules, such as peptides, that can be made synthetically or recombinantly. Member peptides in a combinatorial peptide library include amino acids incorporated randomly into certain or all positions of their sequences. Such libraries have been generated and used in various ways to screen for peptide candidates which bind effectively to target molecules and to identify such sequences.

Many methods for generating peptide libraries have been developed and described. For example, members of the peptide library can be created by split-synthesis performed on a solid support such as polystyrene or polyacrylamide resin, as described by Lam *et al.* (1991) *Nature*

354:82 and PCT publication WO 92/00091. The method disclosed by U.S. Pat. No. 4,833,092 involves the synthesis of peptides in a methodical and predetermined fashion, so that the placement of each library member peptide gives information concerning the synthetic structure of that peptide.

5   Phage display of peptide libraries has become a powerful tool for rapidly screening and identifying novel ligands of virtually any protein target. Of particular interests are display methods using filamentous bacteriophages. U.S. Pat. No. 5,821,047. This method allows the preparation of libraries as large as 10^{10}-10^{12} unique peptide members, many orders of magnitude larger than libraries that may be prepared synthetically. In addition to large library sizes,

10   advantages of phage display include ease of library construction (Kunkel mutagenesis), coupling of the binding entity (displayed peptide) to a unique identifier (its DNA sequence), a selection protocol for amplifying rare binding clones in a pool, and the high fidelity of biosynthesis (compared to synthetic methods). Furthermore, rapid and inexpensive selection protocols are available for identifying those library members that bind to a target of interest. However, only

15   natural peptides composed of L-amino acids may be displayed on phage, so the problem of defining three-dimensional structure-activity relationships is more difficult than it might be for a constrained peptidomimetic containing non-naturally occurring amino acids or nonpeptide components.

      One possible solution to this problem is to use the structural constraints of a folded protein

20   to present small variable peptide segments. Considerable effort has been devoted to introducing structural constraints into combinatorial peptide libraries so that the member peptides represent more closely their native states. Several protein scaffolds capable of presenting a sequence of interest in a conformationally-restricted fashion have been identified, including minibody structures (Bianchi *et al.* (1994) *J Mol Biol* **236**:649-659), β sheets, coiled-coil stem structures

25   (Myszka & Chaiken (1994) *Biochem* **33**:2363-2372), zinc-finger domains, cysteine-linked (disulfide) structures, transglutaminase linked structures, cyclic peptides, helical barrels or bundles, leucine zipper motifs (Martin *et al.* (1994) *EMBO J* **13**:5303-5309), etc.

      A number of identified scaffolds have been used in the construction of combinatorial peptide libraries with structural constraints. U.S. Pat. No. 5,824,483 describes a synthetic peptide

library containing peptides featuring α-helical conformation and thus capable of forming coiled-coil dimers with each other. McBride *et al.* (1996) *J Mol Biol* **259**:819-827 describe a synthetic library of cyclic peptides mimicking the anti-tryptic loop region of an identified proteinase inhibitor. WO 00/20574 and U.S. Pat. No. 6,180,343B1 describe fusion constructs using scaffold proteins such as green fluorescent protein (GFP). Several small protein domains have also been proposed as peptide display scaffolds. Nygren & Uhlen (1997) *Curr. Opin. Struct. Biol.* **7**:463-469; Vita *et al.* (1998) *Biopolymers* **47**:93-100; Vita *et al.* (1999) *Proc. Natl. Acad. Sci. USA* **96**:13091-13096; Smith *et al.* (1998) *J. Mol. Biol.* **277**:317-332; Gururaja et al. (2000) *Chem. & Biol.* **7**:515-527; Christmann *et al.* (1999) *Protein Engng.* **12**:797-806.

Among the identified protein scaffolds, β-turns (hairpins) have been implicated as an important site for molecular recognition in many biologically active peptides. Smith & Pease (1980) *CRC Crit Rev Biochem* **8**:315-300. Thus, peptides containing conformationally-constrained β-turns are particularly desirable. The great majority of the identified β-turn bearing peptides are cyclopeptides which have been generated by the cyclization of a peptide similar to a sequence in the natural substrate. Milner-White (1989) *Trends Pharmacol Sci* **10**:70-74. These cyclopeptides, however, may still retain significant flexibility. For this reason, many studies have attempted to introduce rigid, nonpeptide compounds which mimic the β-turn. Peptides with such nonpeptide β-turn mimic provide useful leads for drug discovery. Ball & Alewood (1990) *J Mol Recog* **3**:55-64; WO 94/03494. The structural mechanisms by which β-turns are stabilized, and specific strand registers are selected, continue to be the subject of considerable interest.

Several examples have been reported of disulfide-constrained peptides intended to mimic protein hairpins or as de novo designed hairpins. In many cases the design includes D-cysteines at one or both ends, as it was initially thought that disulfide bond geometry was not compatible with the cross-strand geometry of hairpins. However, there are some examples that do use L-cys. Evidence for structure is lacking in most studies of disulfide-cyclized peptides. Examples listed here are those whose structures have been experimentally determined, or that use no unusual amino acids and have potency close to a larger, hairpin-containing natural protein in a biological assay.

The structure of a hexapeptide (Boc-CL-Aib-AVC-NMe) was determined crystallographically, revealing a type II' turn and β-sheet geometry. Karle *et al. J. Am. Chem. Soc.* (1988) **110**:1958-1963. An octapeptide with the same cysteine spacing was studied by NMR, and has a similar structure with a turn centered on Pro-Gly. Walse *et al.* (1996) *J.*

5   *Comput.-Aided Mol. Des.* **10**:11-22. Peptides of the form Ac-CXPGXC-NMe were evaluated by measurement of disulfide exchange equilibria, which indicated turn preferences between peptides of as much as 1 kcal/mol. Milburn *et al.* (1987) *J. Am. Chem. Soc.* **109**:4486-4496.

An eleven-residue cyclic peptide, CGVSRQGKPYC, based on the gene 5 protein from M13 is stably structured in aqueous solution, as demonstrated by NMR analysis. The cyclic

10   peptide adopts a structure that is quite similar to the corresponding protein loop. The authors claim that well-defined β-hairpin structure had not been previously reported for any unprotected disulfide-constrained cycle. Rietman *et al.* (1996) *Eur. J. Biochem.* **238**:706-713. This peptide has a Val-Pro pair at the nonhydrogen-bonded sites nearest to the cysteines.

Disulfide-cyclized peptides from the hairpin region of a rabbit defensin have antibacterial

15   activity exceeding (about 5 to 10-fold) that of the linear analogs. Circular dichroism spectroscopy indicates some non-random structure in phosphate buffer. The more potent peptide (CAGFMRIRGRIHPLCMRR) has a Gly-Pro pair at the nonhydrogen-bonded sites nearest to the cysteines. Thennarasu & Nagaraj (1999) *Biochem. Biophys. Res. Commun.* **254**:281-283.

Several peptides from the loops of domain 1 of human CD4 have been studied in Zhang *et*

20   *al.* (1996) *Nature Biotechnology* **14**:472-475; Zhang *et al.* (1997) *Nature Biotechnology* **15**:150-154. In addition to a disulfide constraint, the authors have added exocyclic aromatic amino acids to the peptide termini. No evidence for structure is given, but one cyclic peptide was reported to antagonize both normal CD4 interactions and those involved in CD4-mediated cell entry by HIV.

Few examples exist of small peptides that form a stable tertiary structure without

25   assistance from disulfide bonds or metal ions. Most natural peptides encompassing hairpins are mainly devoid of structure in water or form aggregates. Ramirez-Alvarado et al. (1997) *Protein Sci.* **6**:162-147. A hairpin peptide derived from the B1 domain (the 41-56 residue fragment) of protein G (GB1) has been reported to form a well-populated hairpin (about 50%) in water. Blanco *et al.*(1994) *Nat. Struct. Biol.* **1**:584-590. The GB1 hairpin has four threonine residues at

hydrogen-bonded sites in the strands, including one Thr-Thr cross-strand pair. This is generally believed to be an unfavorable pairing. In addition, there are Trp-Val and Tyr-Phe pairs at adjacent nonhydrogen-bonded sites that might interact to form a small hydrophobic core.

Analysis of hairpin sequences in crystal structures has allowed the *de novo* design of a series of β-hairpin peptides based on the BH8 peptide. Ramirez-Alvarado *et al.* (1996) *Nat. Struct. Biol.* **3**:604-612. The target structure was a type I' turn flanked by three-residue strands. Arg-Gly sequences were added to the ends to improve solubility. One peptide was partially folded into a hairpin conformation (about 30%) as determined by NMR. The importance of inter-strand side chain-side chain interactions was indicated by replacement of certain strand residues with alanine. None of the alanine-substituted peptides showed any tendency to form a hairpin. The same authors reported a second series of experiments in which position i+1 of the turn was varied. Ramirez-Alvarado *et al.* (1997) *J. Mol. Biol.* **273**:898-912. No peptide was more structured than the original sequence with Asn in the turn. A review describing this work suggested that adding Glu-Lys pairs to the termini of the model peptide may help to stabilize the hairpin. Ramirez-Alvarado *et al.* (1999) *Bioorg. Med. Chem.* **7**:93-103.

A peptide comprising the N-terminal 17 residues of the globular protein ubiquitin has been shown to form a native-like hairpin in both aqueous methanol and water, albeit at low apparent population. Zerella et al. (1999) *Protein Sci.* **8**:1320-1331. A recent study, Zerella et al. (2000) *Protein Sci.* **9**:2142-2150, focused on the contributions to the stability of the isolated peptides by residues within the turn region. The data indicated that in a peptide where Thr at position 9 was replaced by Asp, U(1-17)T9D, the native conformation was stabilized significantly over that of the wild type sequence. The estimated population of the folded hairpin was only 64%. Moreover, as the authors noted, the structure of the folded state of U(1-17)T9D may be more dynamic than indicated by the final ensemble. The reason for the greater stability upon substitution of the turn residue remains uncertain.

It is an object of the present invention to provide a simple model system for displaying small peptides with stable hairpin structure and methods of using such a model system in constructing and screening constrained peptide libraries useful in biological and therapeutic applications.

SUMMARY OF THE INVENTION

The present invention is based on the surprising identification of a novel structure motif, the tryptophan zipper (trpzip), that enables the stabilization of hairpin structures in very short peptides. Some of the trpzip peptides showing stable tertiary structures have a minimum length of 10-12 amino acids. Therefore, in one aspect, the invention provides a minimal peptide scaffold having the newly identified stable trpzip motif, comprising a presented turn sequence flanked by two opposite strands with a defined backbone hydrogen-bonding pattern, each strand comprising at least two Trp residues at non-hydrogen-bonded positions. The four Trp residues from the two strands form two Trp-Trp pairs that constitute a cross-strand zipper-like motif with great structural stability. Significantly, the trpzip motif does not require any disulfide bonds.

In one aspect, the presented turn sequence comprises at least 4 amino acids. In another aspect, the presented turn sequence comprises at least 6 amino acids. In addition to the four Trp residues, the two flanking strands comprise other amino acids, preferably naturally occurring L-form amino acids. In one preferred embodiment, the peptide scaffold has a minimum length of 10 amino acids, with 4 amino acids as the presented turn sequence and 3 amino acids each for the flanking strands. In other preferred embodiments, additional residues are included in the strand region and or the turn region of the scaffold. As such, some preferred peptide scaffolds comprise 12, 14, 16, 18 or 20 amino acids. More preferably, the scaffold is no more than 20 amino acids in length.

The invention also encompasses libraries of structurally-constrained peptides, each peptide having the trpzip scaffold as described above, wherein the presented turn sequence consists of random amino acids. Methods of constructing such libraries are also contemplated. The subject libraries can be used for selecting novel peptides capable of binding to identified target molecules. Accordingly, the invention provides methods of identifying peptides capable of binding to a bioactive target molecule, comprising the steps of: a) providing a library of peptides comprising the novel trpzip scaffold; b) contacting the library with the target molecule; c) selecting from the library peptides capable of forming a noncovalent complex with the binding partner; and d) optionally isolating the peptides selected in step c). The selected peptides are useful per se as diagnostics or therapeutics (e.g., agonists or antagonists) used in treatment of

biological organisms. Compositions and methods of the invention may also be useful in analyzing the structure-activity relationship of proteins of interest, thereby providing information for rational drug design.

5   BRIEF DESCRIPTION OF DRAWINGS

Figures 1A-1C are graphs showing the folding of trpzips 1-3. **(1A)** Circular dichroism (CD) spectrum of trpzip1. The near UV region is shown as an inset with a 10-fold expanded y-axis. **(1B)** Thermal denaturation of trpzip1 (20 μM) monitored by CD. The forward melting curve is shown as open circles, while the reverse melting curve is shown as the error bars
10   associated with signal averaging during data acquisition. The first derivatives of melting curves (20, 50, 100, and 150 μM peptide) are overlaid in the inset. **(1C)** Temperature dependence of folding for trpzips 1-3 (calculated from the thermodynamic parameters listed in Table 2).

Figure 2 is a graph depicting the equilibrium ultracentrifugation of trpzips 1-3. The data shown are for 60 μM peptide samples and a rotor speed of 40 krpm. Apparent molecular weights
15   obtained from the slopes (assuming ideal behavior) are shown; calculated formula weights are 1608 for trpzips 1 and 2 and 1648 for trpzip3. Trpzip1 data are offset vertically (ln absorbance - 0.085) for clarity.

Figures 3A-3C depict NMR structures of trpzips 1 and 2. **(3A)** A representative structure of trpzip1 calculated based on NMR-derived restraints. The residues and their positions are indicated. **(3B/3C)** Representative structures of trpzips 1 and 2 aligned on the backbone
20   atoms of residues 2-5 and 8-11 (r.m.s.d. of the mean coordinates of the aligned backbone atoms in the two ensembles is 0.37 Å); the view in 3C is rotated 90° relative to the view in 3B. The backbone carbonyl of residue 6 is indicated to emphasize the difference in turn geometry between the two structures (type II´ for trpzip1 vs. type I´ for trpzip2).

25   **Figure 4** is a graphic representation of the temperature dependence of folding for trpzips 4-6 (calculated from the thermodynamic parameters listed in Table 2). The estimated curve for gb1 was calculated by assuming that mutations in trpzip4 (i.e., those present in trpzips 5 and 6) have independent and additive effects on hairpin stability ($\Delta G_{unf, gb1} = \Delta G_{unf, trpzip5} - \{\Delta G_{unf, trpzip4} - $

$\Delta G_{unf, trpzip6}$) and is nearly identical in shape to previously reported gb1 denaturation curves based on fluorescence or NMR measurements.

Figure 5 compares NMR structures of trpzip4 (light grey) and gb1 protein (dark grey). The trpzip4 structure is a representative structure from the ensemble of 20 structures calculated based on NMR-derived restraints. The backbone atoms of gb1 protein residues 46-52 were superposed on the mean structure of the trpzip4, yielding an r.m.s.d. of 0.67 Å.

Figure 6 is a graphic representation of the temperature dependence of folding for trpzips 4 and 7-9 (calculated from the thermodynamic parameters listed in Table 4).

DETAILED DESCRIPTION OF THE INVENTION

The design of peptides that have well-defined tertiary structures tests our understanding of the principles governing the folding of larger proteins. Short peptides with significant hairpin structure have recently emerged as β–sheet model systems. In a separate study, the inventors of the present invention have discovered that in a disulfide-cyclized β-hairpin peptide, tryptophan was much more stabilizing in a non-hydrogen-bonded (NHB) strand position than other amino acids. Cochran et al. (2001) *J. Am. Chem. Soc.* **123**:625-632. Paired, cross-strand NHB residues in the Cys-cyclized hairpin made roughly independent contributions to stability; thus, a single tryptophan-tryptophan cross-strand pair was shown to be highly stabilizing (and the best NHB residue pair identified). Cochran et al. WO 00/77194; Russell and Cochran (2000) *J. Am. Chem. Soc.* **122**:12600-12601. The present invention provides a novel structural motif, the tryptophan zipper (trpzip), that greatly stabilizes the β–hairpin conformation in short peptides, without any disulfide bonds. As shown in the Examples, peptides having 12 or 16 amino acids in length with different turn sequences are monomeric and fold cooperatively in water. Surprisingly, the folding free energies of the trpzip peptides exceed substantially those of all previously reported β-hairpins and even those of some larger designed proteins. NMR structures of some of the exemplary trpzip peptides revealed exceptionally well-defined β-hairpin conformations stabilized by cross-strand pairs of indole rings. The peptides of the present invention are the smallest peptides to adopt an unique tertiary fold without requiring metal binding, unusual amino acids, or disulfide crosslinks.

Therefore, the present invention provides a novel peptide scaffold for β-turn display. By "scaffold", "peptide scaffold" or "protein scaffold" is meant an amino acid framework useful for presenting a peptide of interest, in a way that the peptide of interest is accessible to other molecules. Preferably, the peptide scaffold has stable, defined tertiary structure, such that the

5   presented peptide adopts a constrained conformation for display.

The term "β-turn" or "β-hairpin", as used herein, refers to an antiparallel β-sheet structure comprising a turn region flanked by two opposite strands with defined backbone hydrogen-bonding pattern. There are several types of hairpins depending on the types of turn, including for example, types I, I', II, and II'. A "presented turn sequence" refers to the central subset region of

10   a β-turn that forms the actual turn structure. As used herein, the term represents a segment with variable amino acid residues that is to be presented in a combinatorial library display. The segment is used to present randomized residues in searching for sequences exhibiting binding affinities to other target molecules of interest. For example, the presented turn sequences can be sequences capable of serving as substrates or inhibitors, being recognized by antibodies, binding

15   to receptors or ligands, or being useful in column affinity chromatography. Using well known methods such as those further described below, such presented sequences can be identified and isolated for further studies and uses.

The term "tryptophan zipper" or "trpzip" refers to a "zipper-like" peptide motif characterized by four tryptophan residues capable of forming two Trp-Trp cross-strand pairs and

20   stabilizing a β-hairpin tertiary structure. The Trp residues within a trpzip are located at non-hydrogen-bonded positions of the opposite strands.

"A defined backbone hydrogen-bonding pattern" as used herein refers to a tertiary structure with defined conformation that is formed and stabilized by interstrand hydrogen bonding participated by amide and or carboxyl moieties of individual strand residues. "Non-

25   hydrogen-bonded positions" or "NHB positions" as used herein refers to strand positions within the hairpin scaffold that do not contribute to and participate in the hydrogen-bonding pattern. See Sibanda et al. (1989) *J. Mol. Biol.* **229**:759-777 for further description of the hydrogen-bonding patterns of β-hairpins and their nomenclature.

The scaffold of the present invention comprises at least two Trp-Trp NHB cross-strand pairs. The combination of at least two Trp-Trp NHB cross-strand pairs greatly stabilizes β-hairpin structures. As further disclosed in the Examples, several trpzip variants having different turn sequences are highly water-soluble, well-structured, and monomeric. High-resolution NMR structures of the peptides show the two cross-strand Trp pairs interdigitating in a zipper-like motif on the surface of the folded peptide. This arrangement of the indole side chains confers unusual spectroscopic properties on the folded molecules, and folding can therefore be monitored readily by changes in circular dichroism (CD) signal. The stabilities of the tryptophan zippers are significantly higher than those reported for other small β-structures. Indeed, on a per-residue basis, the tryptophan zippers have stabilities comparable to much larger native protein domains.

The scaffolds of the invention are non-cysteine constrained. In other words, the scaffolds do not require the involvement of disulfide bridges between strands in order to maintain the stability of the tertiary structures. As such, trpzip peptides of the invention are particularly useful in applications where the disulfide bond formation is either undesirable or unfavorable. For example, the trpzip scaffold can be used for the intracellular display of peptides.

The trpzip peptides of the invention are among the smallest peptides to adopt an unique and stable tertiary fold without requiring metal binding, unusual amino acids, or disulfide bridges. Previous studies have suggested that the minimal size of a stable protein domain without cysteine bridges is approximately 50 amino acids. Nygren and Uhlen (1997) *Curr. Opin. Struct. Biol.* 7:463-469; Privalov and Gill (1988) *Adv. Protein Chem.* 39:191-243. Because of their small size, unusual stability, and very favorable spectroscopic properties, the trpzip scaffolds of the invention provide a useful and simple system for the study and display of β-turns.

The invention provides a library of trpzip peptides for turn display. Preferably, the presented turn sequence consists of random amino acids. Randomization of the turn sequences can be achieved by using methods and techniques well known in the art. Generally, at least 2, preferably at least 4, more preferably at least 6, even more preferably at least 10 amino acid positions need to be randomized. In a preferred embodiment, the random peptide sequence is provided by oligonucleotide synthesis using randomized codon assignments. It should be realized, however, that in a library system encoded by random nucleotides, codons encoding stop

signals (i.e., TAA, TGA and TAG) may be undesirably introduced into the structure. For example, in a synthesis with NNN as the random region, there is a 3/64 chance that the codon will be a stop codon. Thus, in a region of 10 residues, there is a likelihood that 46.7% of the peptides will prematurely terminate. In order to alleviate this problem, random residues can be encoded,

5 for example, as NNK or NNS instead, where K= T or G; and S=C or G. This allows for encoding of all potential amino acid (changing their relative representation slightly), yet preventing the encoding of two stop residues TAA and TGA.

In a preferred embodiment, the peptide library is "fully randomized," meaning there are no sequence preferences or fixed residues at any position within the turn region. In another

10 preferred embodiment, the library is randomized with bias. That is, some positions within the region are either held constant, or selected from a limited number of possibilities. For example, in a preferred embodiment, the residues are randomized within a defined category, such as of hydrophobic residues, hydrophilic residues, aliphatic residues, unbranched residues, branched residues, or aromatic residues, etc. In a preferred embodiment, the random residues are biased to

15 β-turn formation. In addition to random residues in the turn regions, the invention also encompasses amino acid variations at strand positions of the scaffold, other than those occupied by the core Trp residues. For example, variations can occur at NHB strand sites and/or hydrogen-bonded strand sites. A position and its cross-strand pairing partner can have the same or different residues.

20 Many methods for generating peptide libraries are known in the art and can be used to generate the libraries of the invention. In one embodiment, members of the peptide library can be created by split-synthesis performed on a solid support such as polystyrene or polyacrylamide resin, as described by Lam *et al.* (1991) *Nature* 354:82 and PCT publication WO 92/00091. In another embodiment, the trpzip scaffold of the invention can be used in constructing and

25 displaying intracellular peptide libraries.

A preferred method of generating the library of the present invention is phage display. Bacteriophage (phage) display is a known technique by which variant polypeptides are displayed as fusion proteins to the coat protein on the surface of bacteriophage particles (Scott, J.K. and Smith, G. P. (1990) *Science* 249: 386). A "phagemid" is a plasmid vector having a bacterial

origin of replication, e.g., ColE1, and a copy of an intergenic region of a bacteriophage. The phagemid may be based on any known bacteriophage, including filamentous bacteriophage. The plasmid will also generally contain a selectable marker for antibiotic resistance. Segments of DNA cloned into these vectors can be propagated as plasmids. When cells harboring these

5   vectors are provided with all genes necessary for the production of phage particles, the mode of replication of the plasmid changes to rolling circle replication to generate copies of one strand of the plasmid DNA and package phage particles. The phagemid may form infectious or non-infectious phage particles. This term includes phagemids which contain a phage coat protein gene or fragment thereof linked to a heterologous peptide gene as a gene fusion such that the

10   heterologous peptide is displayed on the surface of the phage particle.

The term "coat protein" means a protein, at least a portion of which is present on the surface of the virus particle. From a functional perspective, a coat protein is any protein which associates with a virus particle during the viral assembly process in a host cell, and remains associated with the assembled virus until it infects another host cell. The coat protein may be the

15   major coat protein or may be a minor coat protein. A "major" coat protein is a coat protein which is present in the viral coat at 10 copies of the protein or more. A major coat protein may be present in tens, hundreds or even thousands of copies per virion.

A "fusion protein" is a polypeptide having two portions covalently linked together, where each of the portions is a polypeptide having a different property. The property may be a

20   biological property, such as activity in vitro or in vivo. The property may also be a simple chemical or physical property, such as binding to a target molecule, catalysis of a reaction, etc. The two portions may be linked directly by a single peptide bond or through a peptide linker containing one or more amino acid residues. Generally, the two portions and the linker will be in reading frame with each other.

25   In one preferred embodiment, the trpzip peptides are fused to at least a portion of a phage coat protein to form a fusion protein. The fusion protein can be made by expressing a gene fusion encoding the fusion protein using known techniques of phage display such as those described below. The fusion protein may form part of a phage or phagemid particle in which one or more copies of the trpzip peptide are displayed on the surface of the particle. A gene comprising a

nucleic acid encoding the trpzip peptide or the fusion protein are within the scope of the invention.

In another embodiment, the invention is a method comprising the steps of constructing a library containing a plurality of replicable expression vectors, each expression vector comprising

5 a transcription regulatory element operably linked to a gene fusion encoding a fusion protein, wherein the gene fusion comprises a first gene encoding a trpzip peptide of the invention and a second gene encoding at least a portion of a phage coat protein, where the library comprises a plurality of genes encoding variant trpzip peptide fusion proteins. Variant first genes and libraries thereof encoding variant trpzip peptides are prepared using known mutagenesis techniques

10 described in more detail below.

The invention also includes expression vectors comprising the fusion genes noted above, as well as a library of these vectors. The library of vectors may be in the form of a DNA library, a library of virus (phage or phagemid) particles containing the library of fusion genes or in the form of a library of host cells containing a library of the expression vectors or virus particles.

15 The invention also contemplates a method of selecting novel binding peptides capable of binding to a bioactive target molecule. By "binding peptide" as used herein is meant any peptide that binds with a selectable affinity to a target molecule. By "bioactive target molecule" as used herein is meant any molecule exerting any biological activity in vitro or in vivo, for which it is desirable to produce a ligand. Preferably, the target molecule is a protein. More preferably, the

20 target molecules include receptors, hormone ligands, growth factors, antigens, antibodies, enzymes and enzyme substrates.

In a preferred embodiment, the method of selecting novel binding peptides comprises the steps of: (a) constructing a library of variant replicable expression vectors comprising a transcription regulatory element operably linked to a gene fusion encoding a fusion protein

25 wherein the gene fusion comprises a first gene encoding the trpzip peptide, and a second gene encoding at least a portion of a phage coat protein, where the variant expression vectors comprise variant first genes; (b) transforming suitable host cells with the vectors; (c) culturing the transformed host cells under conditions suitable for forming recombinant phage or phagemid virus particles containing at least a portion of the expression vector and capable of transforming

. the host, so that the particles display one or more copies of the fusion protein on the surface of the particle; (d) contacting the particles with a target molecule so that at least a portion of the particles bind to the target molecule; and (e) separating the particles that bind from those that do not. In the method of the invention, the phage coat protein is preferably the gene III or gene VIII

5 coat protein of a filamentous phage such as M13. Further, preferably the culturing of the transformed host cells is under conditions suitable for forming recombinant phage or phagemid particles where the conditions are adjusted so that no more than a minor amount of phage or phagemid particles display one or more copies of the fusion protein on the surface of the particle (monovalent display).

10 The invention also includes a method of introducing structural bias into a phage-displayed library, using steps (a) through (e) described above. The invention further includes a method of selecting beta-hairpin forming peptide structures from a phage-displayed library, using steps (a) through (e) described above where the target is known to bind beta-hairpin peptide structures, preferably a protein target known to so bind.

15 The utility of phage display lies in the fact that large libraries of selectively randomized protein variants (or randomly cloned cDNAs) can be rapidly and efficiently sorted for those sequences that bind to a target molecule with high affinity. Display of peptide (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci. USA* **87**:6378) or protein (Lowman *et al.* (1991) *Biochemistry* **30**:10832; Clackson *et al.* (1991) *Nature* **352**: 624; Marks *et al.* (1991), *J. Mol. Biol.* **222**:581;

20 Kang *et al.* (1991) *Proc. Natl. Acad. Sci. USA* **88**:8363) libraries on phage have been used for screening millions of polypeptides for ones with specific binding properties (Smith (1991) *Current Opin. Biotechnol.* **2**:668). Sorting phage libraries of random mutants requires a strategy for constructing and propagating a large number of variants, a procedure for affinity purification using the target receptor, and a means of evaluating the results of binding enrichments.

25 Typically, variant polypeptides, such as the trpzip compounds of the invention, are fused to a gene III protein, which is displayed at one end of the virion. Alternatively, the variant polypeptides may be fused to the gene VIII protein, which is the major coat protein of the virion. Such polyvalent display libraries are constructed by replacing the phage gene III with a cDNA encoding the foreign sequence fused to the amino terminus of the gene III protein.

Monovalent phage display is a process in which a protein or peptide sequence is fused to a portion of a gene III protein and expressed at low levels in the presence of wild-type gene III protein so that particles display mostly wild-type gene III protein and one copy or none of the fusion protein (Bass *et al.* (1990) *Proteins* 8:309; Lowman, H.B. and Wells, J.A. (1991) *Methods:*

5   *a Companion to Methods in Enzymology* 3:205). Monovalent display has the advantage over polyvalent phage display that progeny phagemid particles retain full infectivity. Avidity effects are reduced so that sorting is on the basis of intrinsic ligand affinity, and phagemid vectors, which simplify DNA manipulations, are used.   One preferred phage display system is described in U.S. Pat. No. 5,821,047.

10   A two-step approach may be used to select high affinity ligands from peptide libraries displayed on M13 phage. Low affinity leads are first selected from naive, polyvalent libraries displayed on the major coat protein (protein VIII). The low affinity selectants are subsequently transferred to the gene III minor coat protein and matured to high affinity in a monovalent format.

Although most phage display methods have used filamentous phage, other phage display

15   systems, such as lambda phage, T4 phage and T7 phage display systems are also known and can be used to create a library of the trpzip peptides of the invention. WO 95/34683; U.S Pat. No. 5,627,024; Ren *et al.* (1998) *Gene* 215:439; Zhu (1997) *CAN* 33:534; Jiang *et al.* (1997) *CAN* 128:44380; Ren *et al.* (1997) *CAN* 127:215644; Ren (1996) *Protein Sci.* 5:1833; Efimov *et al.* (1995) *Virus Genes* 10:173; Smith & Scott (1993) *Methods in Enzymology* 217:228-257; U. S.

20   Pat. No. 5,766,905.

Suitable gene III vectors for display of trpzip peptides of the invention include fUSE5 (Scott, J. K., and Smith G. P. (1990) *Science* 249:386-390); fAFF1 (Cwirla *et al.* (1990). *Proc. Natl. Acad. Sci. U.S.A.* 87:6378-6382); fd-CAT1 (McCafferty *et al.* (1990) *Nature (London)* 348:552-554); m663 (Fowlkes *et al.* (1992) *Biotechniques* 13:422-427); fdtetDOG, pHEN1

25   (Hoogenboom *et al.* (1991) *Nucleic Acids Res.* 19:4133-4137); pComb3 (Gram *et al.* (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:3576-3580); pCANTAB 5E (Pharmacia); and LamdaSurfZap (Hogrefe (1993)*Gene* 137:85-91).

Phage display methods for proteins, peptides and mutated variants thereof, including constructing a family of variant replicable vectors containing a transcription regulatory element

operably linked to a gene fusion encoding a fusion polypeptide, transforming suitable host cells, culturing the transformed cells to form phage particles which display the fusion polypeptide on the surface of the phage particle, contacting the recombinant phage particles with a target molecule so that at least a portion of the particle bind to the target, separating the particles which

5   bind from those that do not bind, are known and may be used with the method of the invention. See WO 97/29185; O'Boyle *et al.* (1997) *Virology* **236**:338-347; Soumillion *et al.* (1994) *Appl. Biochem. Biotech.* **47**:175-190; O'Neil and Hoess. (1995) *Curr. Opin. Struct. Biol.* **5**:443-449; Makowski (1993) *Gene* **128**:5-11; Dunn (1996) *Curr. Opin. Struct. Biol.* **7**:547-553; Choo and Klug (1995) *Curr. Opin. Struct. Biol.* **6**:431-436; Bradbury & Cattaneo (1995) *TINS* **18**:242-249;

10   Cortese *et al.*, (1995) *Curr. Opin. Struct. Biol.* **6**:73-80; Allen *et al.* (1995) *TIBS* **20**:509-516; Lindquist & Naderi (1995) *FEMS Micro. Rev.* **17**:33-39; Clarkson & Wells (1994) *Tibtech.* **12**:173-184; Barbas (1993) *Curr. Opin. Biol.* **4**:526-530; McGregor (1996) *Mol. Biotech.* **6**:155-162; Cortese *et al.* (1996) *Curr. Opin. Biol.* **7**:616-621; McLafferty *et al.* (1993) *Gene* **128**:29-36.

The gene encoding the coat protein of the phage and the gene encoding the desired trpzip

15   peptide portion of the fusion protein of the invention (i.e., the trpzip peptide of the invention fused to at least a portion of a phage coat protein) can be obtained by methods known in the art. The DNA encoding the gene may be chemically synthesized (Merrifield (1963) *J. Am. Chem. Soc.* **85** :2149) and then mutated to prepare a library of variants as described below.

To ligate DNA fragments together to form a functional vector containing the gene fusion,

20   the ends of the DNA fragments must be compatible with each other. In some cases, the ends will be directly compatible after endonuclease digestion. However, it may be necessary to first convert the sticky ends commonly produced by endonuclease digestion to blunt ends to make them compatible for ligation. To blunt the ends, the DNA is treated in a suitable buffer for at least 15 minutes at 15°C with 10 units of the Klenow fragment of DNA polymerase I (Klenow) in

25   the presence of the four deoxynucleotide triphosphates. The DNA is then purified by phenol-chloroform extraction and ethanol precipitation or other DNA purification technique.

The cleaved DNA fragments may be size-separated and selected using DNA gel electrophoresis. The DNA may be electrophoresed through either an agarose or a polyacrylamide matrix. The selection of the matrix will depend on the size of the DNA fragments to be

separated. After electrophoresis, the DNA is extracted from the matrix by electroelution, or, if low-melting agarose has been used as the matrix, by melting the agarose and extracting the DNA from it, as described in sections 6.30-6.33 of Sambrook *et al.*

5 The DNA fragments that are to be ligated together (previously digested with the appropriate restriction enzymes such that the ends of each fragment to be ligated are compatible) are put in solution in about equimolar amounts. The solution will also contain ATP, ligase buffer and a ligase such as T4 DNA ligase at about 10 units per 0.5 µg of DNA. If the DNA fragment is to be ligated into a vector, the vector is at first linearized by cutting with the appropriate restriction endonuclease(s). The linearized vector is then treated with alkaline phosphatase or calf

10 intestinal phosphatase. The phosphatasing prevents self-ligation of the vector during the ligation step.

After ligation, the vector with the foreign gene now inserted is purified and transformed into a suitable host cell. A preferred transformation method is electroporation. Electroporation may be carried out using methods known in the art. More than one (a plurality) electroporations

15 may be conducted to increase the amount of DNA which is transformed into the host cells. Repeated electroporations are conducted as described in the art. See, for example, Vaughan *et al.* (1996) *Nature Biotechnology* **14**:309-314. The number of additional electroporations may vary as desired from several (2,3,4,...10) up to tens (10, 20, 30,...100) and even hundreds (100, 200, 300,...1000). Repeated electroporations may be desired to increase the size of a combinatorial

20 library, e.g. an antibody library, transformed into the host cells.

Preferably, for library construction, the DNA is present at a concentration of 25 micrograms/ml or greater. More preferably, the DNA is present at a concentration of about 30 micrograms/ml or greater, more preferably at a concentration of about 70 micrograms/ml or greater and even more preferably at a concentration of about 100 micrograms/ml or greater even

25 up to several hundreds of micrograms/ml. Generally, the electroporation will utilize DNA concentrations in the range of about 50 to about 500 micrograms/ml. A time constant during electroporation greater than 3.0 milliseconds (ms) results in a high transformation efficiency.

The DNA is preferably purified to remove contaminants. The DNA may be purified by any known method, however, a preferred purification method is the use of DNA affinity

purification. The purification of DNA, e.g., recombinant plasmid DNA, using DNA binding resins and affinity reagents is well known and any of the known methods can be used in this invention (Vogelstein, B. and Gillespie; D. (1979) *Proc. Natl. Acad. Sci. USA* **76**:615; Callen, W. (1993) *Strategies* **6**:52-53). Commercially available DNA isolation and purification kits are also

5 available from several sources including Stratagene (CLEARCUT Miniprep Kit), and Life Technologies (GLASSMAX DNA Isolation Systems). Non-limiting examples of suitable methods for DNA purification include column chromatography, the use of hydroxylated silical polymers, rehydrated silica gel, boronated silicates, modified glass fiber membranes, fluorinated adsorbents, diatomaceous earth, dialysis, gel polymers and the use of chaotropic compounds with

10 DNA binding reagents, all of which are known and widely used in the art. After purification, the DNA is eluted or otherwise resuspended in water, preferably distilled or deionized water, for use in electroporation at the concentrations of the invention. The use of low salt buffer solutions is also contemplated.

Any suitable cells which can be transformed by electroporation may be used as host cells

15 in the method of the present invention. Suitable host cells which can be transformed include gram negative bacterial cells such as *E. coli*. Suitable *E. coli* strains include JM101, *E. coli* K12 strain 294 (ATCC number 31,446), *E. coli* strain W3110 (ATCC number 27,325), *E. coli* X1776 (ATCC number 31,537), *E. coli* XL-1Blue (Stratagene), and *E. coli* B; however many other strains of *E. coli*, such as XL1-Blue MRF', SURE, ABLE C, ABLE K, WM1100, MC1061,

20 HB101, CJ136, MV1190, JS4, JS5, NM522, NM538, and NM539, may be used as well. Cells are made competent using known procedures.

Cell concentrations of about 10^{10} colony forming units (cfu)/mL) of viable living cells and greater are preferably used for electroporation. More preferably, the viable cells are concentrated to about 1×10^{11} to about 4×10^{11} cfu/mL. Preferred cells which may be

25 concentrated to this range are the SS320 cells described below. Cells are preferably grown in culture in standard culture broth, optionally for about 6-48 hrs (or to $OD_{600} = 0.6 - 0.8$) at about 37°C, and then the broth is centrifuged and the supernatant removed (e.g. decanted). Initial purification is preferably by resuspending the cell pellet in a buffer solution (e.g. HEPES pH 7.4) followed by recentrifugation and removal of supernatant. The resulting cell pellet is resuspended

30 in dilute glycerol (e.g. 5 – 20% v/v) and again centrifuged to form a cell pellet and the

18

supernatant removed. The final cell concentration is obtained by resuspending the cell pellet in water or dilute glycerol to the desired concentration.

A particularly preferred recipient cell for the electroporation is a competent *E. coli* strain containing a phage F′ episome. Any F′ episome which enables phage replication in the strain

5   may be used in the invention. Suitable episomes are available from strains deposited with ATCC or are commercially available (CJ236, CSH18, DH5alphaF', JM101, JM103, JM105, JM107, JM109, JM110), KS1000, XL1-BLUE, 71-18 and others). Strain SS320 was prepared by mating MC1061 cells with XL1-BLUE cells under conditions sufficient to transfer the fertility episome (F′ plasmid) of XL1-BLUE into the MC1061 cells. In general, mixing cultures of the two cell

10 types and growing the mixture in culture medium for about one hour at 37°C is sufficient to allow mating and episome transfer to occur. The new resulting *E. coli* strain has the genotype of MC1061 which carries a streptomycin resistance chromosomal marker and the genotype of the F′ plasmid which confers tetracycline resistance. The progeny of this mating is resistant to both antibiotics and can be selectively grown in the presence of streptomycin and tetracycline. Strain

15 SS320 has been deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Virginia, USA on June 18, 1998 and assigned Deposit Accession No. 98795.

This deposit of strain SS320 was made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure

20 and the Regulations thereunder (Budapest Treaty). This assures maintenance of a viable culture for 30 years from the date of deposit. The organisms will be made available by ATCC under the terms of the Budapest Treaty, and subject to an agreement between Genentech, Inc. and ATCC, which assures permanent and unrestricted availability of the progeny of the cultures to the public upon issuance of the pertinent U.S. patent or upon laying open to the public of any U.S. or foreign

25 patent application, whichever comes first, and assures availability of the progeny to one determined by the U.S. Commissioner of Patents and Trademarks to be entitled thereto according to 35 USC §122 and the Commissioner's rules pursuant thereto (including 37 CFR §1.14 with particular reference to 886 OG 638).

The assignee of the present application has agreed that if the cultures on deposit should die or be lost or destroyed when cultivated under suitable conditions, they will be promptly replaced on notification with a viable specimen of the same culture. Availability of the deposited cultures is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

A useful method for identification of certain residues or regions of the peptide that are preferred locations for mutagenesis is called "alanine scanning mutagenesis" as described by Cunningham and Wells (1989) *Science* 244:1081-1085 . Here, a residue or group of target residues are identified (*e.g.*, charged residues such as arg, asp, his, lys, and glu) and replaced by a neutral or negatively charged amino acid (most preferably alanine or polyalanine) to affect the interaction of the amino acids with target molecule. Those amino acid locations demonstrating functional sensitivity to the substitutions then are refined by introducing further or other variants at, or for, the sites of substitution. Thus, while the site for introducing an amino acid sequence variation is predetermined, the nature of the mutation *per se* need not be predetermined. For example, to analyze the performance of a mutation at a given site, ala scanning or random mutagenesis is conducted at the target codon or region and the expressed peptides are screened for the desired activity.

Oligonucleotide-mediated mutagenesis is a preferred method for preparing the substitution, deletion, and insertion variants of the invention. This technique is well known in the art as described by Zoller *et al.* (1987) *Nucleic Acids Res.* **10**: 6487-6504. Briefly, a gene encoding a protein fusion or heterologous polypeptide is altered by hybridizing an oligonucleotide encoding the desired mutation to a DNA template, where the template is the single-stranded form of the plasmid containing the unaltered or native DNA sequence of the gene. After hybridization, a DNA polymerase is used to synthesize an entire second complementary strand of the template which will thus incorporate the oligonucleotide primer, and will code for the selected alteration in the gene. Generally, oligonucleotides of at least 25 nucleotides in length are used. An optimal oligonucleotide will have 12 to 15 nucleotides that are completely complementary to the template on either side of the nucleotide(s) coding for the mutation. This ensures that the oligonucleotide will hybridize properly to the single-stranded DNA template

molecule. The oligonucleotides are readily synthesized using techniques known in the art such as that described by Crea *et al.* (1978) *Proc. Nat'l. Acad. Sci. USA* **75**: 5765.

5   The DNA template is generated by those vectors that are derived from the bacteriophage used in the phage display system, e.g. bacteriophage M13 vectors (the commercially available M13mp18 and M13mp19 vectors are suitable), or those vectors that contain a single-stranded phage origin of replication; examples are described by Viera *et al.* (1987) *Meth. Enzymol.* **153**:3. Thus, the DNA that is to be mutated can be inserted into one of these vectors in order to generate single-stranded template.

10   To alter the native DNA sequence, the oligonucleotide is hybridized to the single stranded template under suitable hybridization conditions. A DNA polymerizing enzyme, usually T7 DNA polymerase or the Klenow fragment of DNA polymerase I, is then added to synthesize the complementary strand of the template using the oligonucleotide as a primer for synthesis. A heteroduplex molecule is thus formed such that one strand of DNA encodes the mutated form of the gene, and the other strand (the original template) encodes the native, unaltered sequence of the gene. This heteroduplex molecule is then transformed into a suitable host cell, usually a prokaryote such as *E. Coli* JM101. After growing the cells, they are plated onto agarose plates and screened using the oligonucleotide primer radiolabelled with 32-Phosphate to identify the bacterial colonies that contain the mutated DNA.

20   The method described immediately above may be modified such that a homoduplex molecule is created wherein both strands of the plasmid contain the mutation(s). The modifications are as follows: The single-stranded oligonucleotide is annealed to the single-stranded template as described above. A mixture of three deoxyribonucleotides, deoxyriboadenosine (dATP), deoxyriboguanosine (dGTP), and deoxyribothymidine (dTTP), is combined with a modified thio-deoxyribocytosine called dCTP-(aS) (which can be obtained from 25   Amersham). This mixture is added to the template-oligonucleotide complex. Upon addition of DNA polymerase to this mixture, a strand of DNA identical to the template except for the mutated bases is generated. In addition, this new strand of DNA will contain dCTP-(aS) instead of dCTP, which serves to protect it from restriction endonuclease digestion. After the template strand of the double-stranded heteroduplex is nicked with an appropriate restriction enzyme, the

template strand can be digested with ExoIII nuclease or another appropriate nuclease past the region that contains the site(s) to be mutagenized. The reaction is then stopped to leave a molecule that is only partially single-stranded. A complete double-stranded DNA homoduplex is then formed using DNA polymerase in the presence of all four deoxyribonucleotide

5 triphosphates, ATP, and DNA ligase. This homoduplex molecule can then be transformed into a suitable host cell such as *E. coli* JM101, as described above.

Mutants with more than one amino acid to be substituted may be generated in one of several ways. If the amino acids are located close together in the polypeptide chain, they may be mutated simultaneously using one oligonucleotide that codes for all of the desired amino acid

10 substitutions. If, however, the amino acids are located some distance from each other (separated by more than about ten amino acids), it is more difficult to generate a single oligonucleotide that encodes all of the desired changes. Instead, other alternative methods may be employed.

In the first method, a separate oligonucleotide is generated for each amino acid to be substituted. The oligonucleotides are then annealed to the single-stranded template DNA

15 simultaneously, and the second strand of DNA that is synthesized from the template will encode all of the desired amino acid substitutions. The alternative method involves two or more rounds of mutagenesis to produce the desired mutant. The first round is as described for the single mutants: wild-type DNA is used for the template, an oligonucleotide encoding the first desired amino acid substitution(s) is annealed to this template, and the heteroduplex DNA molecule is

20 then generated. The second round of mutagenesis utilizes the mutated DNA produced in the first round of mutagenesis as the template. Thus, this template already contains one or more mutations. The oligonucleotide encoding the additional desired amino acid substitution(s) is then annealed to this template, and the resulting strand of DNA now encodes mutations from both the first and second rounds of mutagenesis. This resultant DNA can be used as a template in a third

25 round of mutagenesis, and so on.

Cassette mutagenesis is also a preferred method for preparing the substitution, deletion, and insertion variants of the invention. The method is based on that described by Wells *et al.* (1985) *Gene* **34**:315. The starting material is a plasmid (or other vector) containing the gene to be mutated. The codon (s) in the gene to be mutated are identified. There must be a unique

restriction endonuclease site on each side of the identified mutation site(s). If no such restriction sites exist, they may be generated using the above-described oligonucleotide-mediated mutagenesis method to introduce them at appropriate locations in the gene. After the restriction sites have been introduced into the plasmid, the plasmid is cut at these sites to linearize it. A

5   double-stranded oligonucleotide encoding the sequence of the DNA between the restriction sites but containing the desired mutation(s) is synthesized using standard procedures. The two strands are synthesized separately and then hybridized together using standard techniques. This double-stranded oligonucleotide is referred to as the cassette. This cassette is designed to have 3' and 5' ends that are compatible with the ends of the linearized plasmid, such that it can be directly

10   ligated to the plasmid. This plasmid now contains the mutated DNA sequence of the gene. Vectors containing the mutated variants can be transformed into suitable host cells as described above.

PCR mutagenesis is also suitable for making amino acid sequence variants of the starting polypeptide. See Higuchi, in *PCR Protocols*, pp.177-183 (Academic Press, 1990); and Vallette

15   *et al.*, *Nuc. Acids Res.* 17:723-733 (1989). Briefly, when small amounts of template DNA are used as starting material in a PCR, primers that differ slightly in sequence from the corresponding region in a template DNA can be used to generate relatively large quantities of a specific DNA fragment that differs from the template sequence only at the positions where the primers differ from the template.

20   The transformed cells are generally selected by growth on an antibiotic, commonly tetracycline (tet) or ampicillin (amp), to which they are rendered resistant due to the presence of tet and/or amp resistance genes in the vector.

Suitable phage and phagemid vectors for use in this invention include all known vectors for phage display. Additional examples include pComb8 (Gram *et al.* (1992) *Proc. Natl. Acad.*

25   *Sci. USA* **89**:3576-3580); pC89 (Felici *et al.* (1991) *J. Mol. Biol.* **222**:310-310); pIF4 (Bianchi *et al.* (1995) *J. Mol. Biol.* **247**:154-160); PM48, PM52, and PM54 (Iannolo. (1995) *J. Mol. Biol* **248**:835-844); fdH (Greenwood *et al.* (1991) *J. Mol. Biol.* **220**:821-827); pfd8SHU, pfd8SU, pfd8SY, and fdISPLAY8 (Malik & Perham (1996) *Gene* **171**:49-51); "88" (Smith (1993) *Gene* **128**:1-2); f88.4 (Zhong *et al.* (1994) *J. Biol. Chem*, **269**:24183-24188); p8V5 (Affymax); MB1,

MB20, MB26, MB27, MB28, MB42, MB48, MB49, MB56: (Markland *et al.* (1991) *Gene* **109**:13-19). Similarly, any known helper phage may be used when a phagemid vector is employed in the phage display system. Examples of suitable helper phage include M13-KO7 (Pharmacia), M13-VCS (Stratagene), and R408 (Stratagene).

5   After selection of the transformed cells, these cells are grown in culture and the vector DNA may then be isolated. Phage or phagemid vector DNA can be isolated using methods known in the art, for example, as described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd edition, (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

        The isolated DNA can be purified by methods known in the art. This purified DNA can
10   then be analyzed by DNA sequencing. DNA sequencing may be performed by the method of Messing *et al.* (1981) *Nucleic Acids Res.* **9**:309, the method of Maxam *et al.* (1980) *Meth. Enzymol.* **65**:499, or by any other known method.

        Various aspects and embodiments of the present invention demonstrate the advantages of a novel model system for rationally designing and analyzing peptides of defined structural
15   features. The combinatorial libraries comprising such peptides and methods of using thereof provide useful information and tools for exploring the basic structure-activity relationships involved in almost all biological molecular interactions. The peptides disclosed herein or generated according to the disclosure of the invention can be candidates for various biological or therapeutic agents, including but not limited to, enzyme inhibitors, ligand antagonists, ligand
20   agonists, toxins, and immunogens.

        In one aspect, the trpzip scaffold is used to present random peptide sequences to potential target molecules. Target molecules can be at least a portion of any molecules, including any known or unknown peptides, proteins, other macromolecules or chemical compounds that are capable of binding to the peptides and optionally exerting bioactivities. Protein molecules such as
25   receptors, ligands, antigens, antibodies, enzymes, enzyme substrates and inhibitors, and fragments or portions thereof are encompassed by "target molecules." Other non-protein chemical compounds, organic or inorganic, can also be the target molecules of the peptides.

In another aspect, the sequence of an identified trpzip peptide is used to generate more candidate peptides. For example, the sequence may be the basis of subsequent round(s) of (biased) randomization, to develop peptides with desired activities. Alternatively, the identified sequence of the randomized region can be introduced into other peptide scaffold structures to obtain a display with different conformation/shape.

In one aspect, the system provided herein is used to screen for target molecules that bind to the random trpzip peptides. Furthermore, the trpzip peptides that bind to a target molecule with desirable bioactivities can be used to mimic or antagonize the functions of wild type ligand(s) of the identified target molecule.

The trpzip peptides or their binding partner molecules can also be used to generate antibodies for diagnostic and/or therapeutic uses. Methods of making antibodies to identified polypeptides and proteins are known in the art.

The following examples are provided by way of illustration and not by way of limitation. All disclosures of the references cited herein are expressly incorporated herein by reference in their entirety.

EXAMPLES

EXAMPLE 1: DESIGN AND CHARACTERIZATION OF TRPZIP1 AND TURN VARIANTS THEREOF

Methods

Peptide Synthesis

For all the examples described herein below, peptides were synthesized as C-terminal amides using standard Fmoc chemistry on a Pioneer synthesizer (PE Biosystems). Synthesized peptides were cleaved from resin by treatment with 5% triisopropylsilane in trifluoroacetic acid (TFA) for 1.5-4 hours at room temperature. After removal of TFA by rotary evaporation, peptides were precipitated by addition of ethyl ether and then purified by reversed-phase HPLC (acetonitrile/H_2O/0.1% TFA). Peptide identity was confirmed by electrospray mass spectrometry.

CD Spectroscopy and Analysis of Thermal Denaturation Curves

Spectra were acquired with an Aviv Instruments, Inc. Model 202 spectrophotometer. Peptide concentrations were determined spectrophotometrically as described in Gill & von Hippel (1989) *Anal. Biochem.* 182:319-326. Melting curves were acquired at 229 nm with 1.5 min
5   equilibration at each temperature and an averaging time of 15 s. Thermal denaturation was reversible, as judged by recovery of CD signal ($\geq$ 95%) upon cooling. In addition, reverse melting curves were acquired for trpzips 1 and 4. Reverse and forward curves were identical in shape, with ≤ 0.5 K shift in T_m. As a model for the unfolded state of the peptides, the melting curve (linear) of an equimolar mixture of the trpzip1 half peptides SWTWEG and NKWTWK
10   was measured. Data for the trpzip peptides were then fit to a two-state unfolding equilibrium as described in Minor & Kim (1994) *Nature* 367:660-663, fixing the unfolded baseline. Folded baselines, T_m, ΔH_m (ΔH at T_m), and ΔC_p were allowed to vary. For trpzips 5 and 6, the unfolded baseline could be fit directly to the experimental data. ΔS_m was calculated from the fit parameters ($\Delta H_m / T_m$). Errors in Table 2 were generated by the fitting algorithm (Kaleidagraph, Synergy
15   Software) and were given to indicate the quality of the fits to the particular experimental data set. However, when fitting different data sets, ΔH_m and ΔC_p values varied by ~ 10%, as is typical in thermal denaturation experiments. Becktel & Schellman (1987) *Biopolymers* 26:1859-1877

Fitting with ΔC_p fixed to 0 Muñoz et al. (1997) *Nature* 390:196-199; Honda et al. (2000) *J. Mol. Biol.* 295:269-278) resulted in significant overestimates of hairpin population at lower
20   temperatures; this portion of the stability curve was especially sensitive to errors in ΔC_p. Fitting the trpzip denaturation curves in this manner required large shifts in T_m (~5-10 K higher than the minimum in a derivative plot) and generated fits of lower quality. In addition, van't Hoff plots showed clear curvature through the transition region, indicating a non-zero ΔC_p. From our data on trpzips 4, 5 and 6, we can estimate ΔC_p ~200 cal mol^{-1} K^{-1} for the gb1 peptide, which is
25   sufficient to explain the discrepancy between our population estimate for the gb1 hairpin and those previously reported (Munoz et al. (1997), *supra*; Honda et al. (2000), *supra*). Recently, a non-zero ΔC_p (~ 100 cal mol^{-1} K^{-1}) was reported for the unfolding of a 12-residue hairpin related to gb1. Espinosa & Gellman (2000) *Angew. Chem. Int. Ed.* 39:2330-2333.

Analytical Ultracentrifugation

Samples (in 20 mM potassium phosphate, 150 mM KCl, pH 7.1; 277 K) were analyzed in a

Beckman XL-A ultracentrifuge at rotor speeds of 40 and 55 krpm. Peptide concentration was

monitored by absorbance at or near 290 nm. Data for both speeds and two initial peptide

5   concentrations (60 and 200 µM, 11 data sets total per peptide) were fit simultaneously to a

nonideal single species model using the program NONLIN. Johnson et al. (1981) *Biophys. J.* **36**:

575-588.

Allowing nonideality improved the fit for the 200 µM samples while only slightly

changing the reduced apparent molecular weight σ (~ +6%). For all 3 peptides, data from 60 µM

10   samples fit an ideal model (Fig. 2) with random residuals. Expected σ values were determined

from partial specific volumes based on residue composition, calculated buffer density, and

monomer formula weights.

NMR Spectroscopy and Structure Calculations

NMR samples contained 1-3 mM peptide in 92% H_2O/8% D_2O, pH 5.5 (trpzip1 and

15   trpzip2) or pH 6.0 (trpzip4 and gb1, 41-56), with 0.1 mM DSS as a chemical shift reference. All

spectra were acquired on a Bruker DRX-500 or a Varian Unity-400 spectrometer at 15°C. 2QF-

COSY, TOCSY and ROESY spectra were acquired using gradient coherence selection or

excitation sculpting for water suppression, as described in, for example, Cochran et al. (2001),

supra, and references cited therein. Proton resonances were assigned by standard methods.

20   Cochran et al. (2001), *supra*. $^3J_{H^N\text{-}H^\alpha}$ were obtained by fitting Lorentzian lines to the antiphase

doublets of $H^N\text{-}H^\alpha$ peaks in 2QF-COSY spectra processed to high digital resolution in F_2.

$^3J_{H^N\text{-}H^\alpha}$ were extracted from COSY-35 spectra acquired on D_2O solutions of the peptides.

Distance and dihedral angle restraints were generated as described in Skelton *et al.* (1994)

Biochemistry **33**:13581-13592. 80 initial structures were calculated using the hybrid distance

25   geometry/simulated annealing program DGII (Havel *et al.* (1991) *Prog. Biophys. Mol. Biol.*

56:43-78.); 50 of these were further refined by restrained molecular dynamics using the AMBER

all-atom forcefield implemented in DISCOVER as described previously (Skelton *et al.* (1994),

supra). 20 structures having the lowest restraint violation energy and good geometry were chosen to represent the solution conformation of each peptide. The structure with the lowest r.m.s.d. to the average coordinates of the ensemble was chosen as the representative structure.

The concentration-dependence of the NMR spectra of trpzip2 and trpzip4 were evaluated
5   by 1D ^{1}H NMR (10-fold and 100-fold dilution of samples used to acquire 2D data; final concentrations: 1.2 mM, 120 µM, and 12 µM for trpzip2, and 3.2 mM, 320 µM, and 32 µM for trpzip4). For both peptides, there were small chemical shift changes (in all cases $\Delta\delta \leq 0.08$ ppm between concentrated and 10-fold diluted samples, and $\Delta\delta \leq 0.02$ ppm between 10-fold and 100-fold diluted samples). For example, the trpzip2 peak with the largest $\Delta\delta$ was that from W4$^{H\varepsilon3}$; in
10   the 1.2 mM sample this proton resonates at 5.656 ppm (2.0 ppm upfield from the expected random coil position). This peak shifts 0.043 ppm downfield (120 µM sample), and an additional 0.004 ppm downfield upon further dilution (12 µM sample). In contrast, analytical ultracentrifugation indicates that trpzip2 is monomeric up to at least 200 µM. Taken together, these data imply that limited self-association may be occurring at millimolar concentrations. The
15   fact that the $\Delta\delta$ are extremely small indicates that self-association does not significantly perturb the peptide structure; furthermore, there are no NOEs indicative of a specific interaction between monomers. Thus, the calculated structures accurately represent the monomer conformations.

Results

The peptide trpzip1 (Table 1) consists of a representative type II′ turn sequence (EGNK)
20   flanked by the sequence WTW. An additional residue was added to each end of the peptide to permit cross-strand hydrogen bonding between the termini. Residues in hydrogen-bonded positions of the strands were taken from sequences used in our previous studies (WO 00/77194). Surprisingly, given that one-third of the residues are tryptophan, the peptide is freely soluble in water at millimolar concentrations. Trpzip1 has an unusual CD spectrum with intense exciton
25   coupled bands at 215 and 229 nm (Fig. 1A), indicating interaction between the aromatic chromophores. Furthermore, the near UV CD spectrum of trpzip1 has well defined bands at the longer wavelength absorption maxima of tryptophan (Fig. 1A, inset), indicating that the indole side chains are in a defined chiral environment. In proteins, such near UV CD bands are often taken as evidence for fixed tertiary structure.

Trpzip1 has a reversible, cooperative thermal denaturation curve with a midpoint at 323 K (Fig. 1B). The data are of exceptionally high quality for a β-peptide: folding may be monitored sensitively at the 229 nm exciton coupled band, where sample absorbance causes few problems. Very poor signal-to-noise ratio is frequently a problem in CD-monitored folding studies of other small β-structures; see, for example, Kortemme et al. (1998) *Science* **281**: 253-256. Reverse and forward melting curves overlay closely (Fig. 1B), demonstrating that the thermal transition is reversible. The melting temperature does not shift with peptide concentration (20-150 μM; Fig. 1B, inset), suggesting that trpzip1 does not self-associate at these concentrations. The thermal denaturation data fit well to a two-state model and reveal that folding is enthalpically favorable at ambient temperatures, with a significant heat capacity change (Table 2).

Two variants were synthesized, in which the Gly-Asn turn sequence of trpzip1 was replaced by stronger turn promoting sequences (trpzips 2 and 3; Table 1). Trpzip2 and trpzip3 have CD spectra that overlay closely with that of trpzip1 (not shown) and, likewise, exhibit reversible and cooperative melting behavior. Thermodynamic parameters for trpzips 2 and 3 are similar to those of trpzip1, with stability curves (and T_m) shifted to higher temperatures (Fig. 1C; Table 2). Interestingly, the denaturation curve for trpzip2 (Asn-Gly turn) is distinctly more cooperative than those of trpzips 1 or 3 (D-Pro-Asn turn). Trpzip2 also appears to be more stable than trpzip3 at low temperatures, despite previous conclusions that the D-Pro-Asn turn (and the related II′ turn D-Pro-Gly) are more stabilizing than Asn-Gly. Cochran et al. (2001) *J. Am. Chem. Soc.* **123**:625-632; Stanger & Gellman (1998) *J. Am. Chem. Soc.* **120**:4236-4237; Syud et al. (1999) *J. Am. Chem. Soc.* **121**:11577-11578. Instead, the conformational restriction of the D-proline appears to confer additional stability only at relatively high temperatures. Equilibrium ultracentrifugation confirms that all three trpzip peptides sediment as single species of the expected monomer molecular weights (Fig. 2; Table 2).

The three-dimensional structures of trpzip1 and trpzip2 were determined by NMR. All ^{1}H resonances were assigned by conventional 2D methods at 288K, pH 5.5. Resonance assignments and coupling constants for trpzips 1 and 2 are shown in Tables 5 and 6, respectively. 1D data are consistent with the peptides being predominantly monomeric at the millimolar concentrations used to acquire the 2D data; see above in Methods. Overall, the NMR data are of unusually high quality for short, linear peptides and provide strong evidence that the molecules are highly

structured. The chemical shift dispersion is remarkable, allowing accurate measurement of the majority of H^N-H^α and H^α-H^β coupling constants and unambiguous assignment of nearly all NOE peaks; the number and intensity of observed NOE peaks are comparable to those routinely seen with small, stable proteins. Likewise, in addition to NOE-based distance restraints,

5   numerous backbone dihedral angle restraints (derived from extreme $^3J_{H^N-H^\alpha}$) could be included in the structure calculations. Furthermore, the tryptophan sidechain conformations are all well defined, having χ_1 angles of ~ –60° (indicated by analysis of $^3J_{H^\alpha-H^\beta}$ and local ROEs). The high population of the folded state under the conditions of the NMR experiments, as well as the quality of the data, validate the high precision of the structures calculated for these peptides. Trpzip1

10   adopts a β-hairpin conformation with the expected type II′ β-turn (Fig. 3A). Cross-strand tryptophan rings pack intimately against one another, with less contact between adjacent tryptophan pairs. Analysis of trpzip2 reveals a very similar structure, only deviating from trpzip1 by having a type I′ β-turn at residues 6 and 7 (Figs. 3B/3C; Table 3).

EXAMPLE 2: DESIGN AND CHARACTERIZATION OF GB1 VARIANTS CONTAINING THE TRPZIP MOTIF

15   The trpzip peptides may be compared to a previously described β-hairpin peptide taken from the B1 IgG-binding domain of protein G. The peptide gb1 (residues 41-56 of the B1 domain) exhibits partial hairpin character, estimated at ~ 40% (278 K) by NMR. Blanco et al. (1994) *Nature Struct. Biol.* 1:584-590. More recently, the estimated hairpin population has

20   doubled, based on fluorescence-monitored folding studies and additional NMR experiments. Muñoz et al. (1997) *Nature* 390:196-199; Honda et al. (2000) *J. Mol. Biol.* 295: 269-278. The peptide appears to be stabilized by a cluster of four hydrophobic residues (W43, Y45, F52, and V54). From NOEs observed for the peptide, and from the structure of the sequence in the parent protein, the hydrophobic strand residues are expected to occupy adjacent non-hydrogen-bonded

25   sites on one face of the hairpin. This is precisely the arrangement of tryptophan residues in the trpzip peptides, allowing the direct comparison of the gb1 and trpzip hydrophobic clusters.

As expected from the stability of trpzips 1-3, replacement of gb1 residues Y45, F52, and V54 with tryptophan yields an exceptionally well-folded β-hairpin (trpzip4; Table 1). The

thermal melting curve for trpzip4 is more cooperative than those of trpzips 1-3, yielding

thermodynamic parameters that reflect this difference (Table 2). Trpzip4 is also more stable than

trpzips 1-3 at low temperatures, resulting in a modest increase in folded population (Figs. 1C and

4). Most importantly, the thermal denaturation curve of trpzip4 is much more cooperative than

5   that of the wt gb1 peptide, and the melting temperature of trpzip4 is higher by *at least* 40 K,

depending on the method used to estimate the folded population of gb1 (Fig. 4).

In contrast, when tryptophan residues 4, 9, and 11 of trpzip1 are replaced with the

appropriate gb1 residues (Y, F, and V, respectively), we find no evidence by NMR for the hairpin

conformation (all $^3J_{H^N-H^\alpha} < 8$ Hz, not shown). This shows that the gb1 hydrophobic cluster is not

10   sufficient to maintain a significant hairpin population without additional stabilizing elements.

To explore this in more detail, we reintroduced individually into trpzip4 the Phe-Tyr and

Trp-Val cross-strand pairs of gb1 (trpzip5 and trpzip6, respectively). Unlike gb1, trpzips 5 and 6

each have one Trp-Trp cross-strand pair, so folding can be monitored by CD (229 nm, as for the

other trpzips). We find both trpzip5 and trpzip6 to be much less stably folded than trpzip4 (Table

15   2; Fig. 4) but more stably folded than wildtype gb1. From our earlier studies in disulfide-cyclized

hairpins, we expect ~ 1 kcal mol^{-1} loss in stability for each gb1 cross-strand pair (Tyr-Phe or Trp-

Val) relative to Trp-Trp. Russell & Cochran (2000) *J. Am. Chem. Soc.* **122**:12600-12601. In

agreement with this expectation, unfolding free energies (298 K) are 1.69, 0.57, and 0.49 kcal

mol^{-1} for trpzips 4, 5, and 6, respectively. Therefore, assuming additive stabilization from the

20   two pairs, we estimate $\Delta G_{unf} \sim -0.6$ kcal mol^{-1} for gb1 at 298 K. Our population estimate for gb1

agrees closely with the lower estimate originally reported. Blanco et al. (1994) *Nature Struct.*

Biol. **1**:584-590.

As observed for the other trpzip peptides, the NMR data for trpzip4 are of exceptional

quality and support the conclusion that the molecule is well folded (Table 7). The fingerprint

25   region of the trpzip4 COSY spectrum shows dramatic chemical shift dispersion, especially when

compared to the spectrum for wildtype gb1 peptide (data not shown). Chemical shifts represent

population-weighted averages of all conformations adopted in solution; therefore, the extreme

H^N and H$^\alpha$ shifts of trpzip4 indicate that the folded conformation is highly populated. From these

data, taken together with the thermal denaturation curves (Fig. 4), we conclude that trpzip4 has a

much higher folded population than the gb1 peptide and that the cross-strand tryptophan pairs of the trpzip motif are superior to the hydrophobic cluster of gb1.

The structure of trpzip4 shows Trp-Trp packing and strand orientations similar to those observed in trpzips 1 and 2 (Table 3; Fig. 5), despite the fact that there are six rather than four

5   intervening turn residues. Trpzip4 extends the strands by another residue and presents a type I β-turn, with K50 adopting a positive φ angle. The turn geometry of trpzip4 is indistinguishable from that of the same turn in the full length B1 domain within the error of the structure determinations (Fig. 5). The twist of the two strands, however, is markedly different between the peptide and the protein; the protein is only modestly twisted (2GB1; Θ ~ 20°), whereas trpzip4 is

10   highly twisted (Θ ~ 70°). This large twist is within the range observed in natural proteins and still allows good hydrogen-bonding geometry. The high degree of twist would appear to result from the cross-strand Trp-Trp packing, since it is observed in all three trpzip structures. The backbone coupling constants for tryptophan residues in the three peptides (7.1-8.2 Hz) are lower than those of the intervening hydrogen-bonded threonine residues (8.9-9.8 Hz), consistent with the

15   alternating less and more negative φ angles that are a hallmark of a twisted sheet. Chothia (1983) *J. Mol. Biol.* **163**:107-117. The geometry of the tryptophan zipper is that expected for an antiparallel β-coiled coil.

EXAMPLE 3: DESIGN AND CHARACTERIZATION OF TRPZIP4 VARIANTS WITH IMPROVED STABILITIES

20   Trpzips 1-6 described above consist of the core strand motif WTW paired with WTW on the opposite strand. Residues other than threonine may be possible at the hydrogen bonding sites in between the two Trp residues of each strand. To explore this, trpzips 7-9 (Table 1) were synthesized, in which the two threonines of trpzip4 are replaced by His-Val, Val-His, and Val-Val pairs, respectively. Trpzips 7-9 formed hairpin structures with CD spectra extremely similar to

25   that shown in Fig 1A for trpzip1. In addition, trpzips 7-9 are all more stable than trpzip4, as determined from thermal denaturation experiments (Table 4), demonstrating that these substitutions for threonine are fully compatible with the trpzip scaffold. Other similar residue substitutions are expected to be compatible as well, for example, Ile instead of Val; and Phe, Tyr, or Trp instead of His.

In conclusion, the trpzip peptides provided herein behave as folded proteins by generally accepted criteria. Presently, they are the smallest all-natural linear polypeptides having such folding behaviors. Their per-residue thermodynamic parameters (ΔG, ΔH, and ΔC_p) are comparable to those of larger protein domains, indicating that, like other proteins, the folding of

5   the trpzip hairpins is driven by burial of hydrophobic surface area (i.e., tryptophan sidechains). Alexander et al. (1992) *Biochemistry* **31**:3597-3603; Becktel & Schellman (1987) *Biopolymers* **26**:1859-1877.

Table 1. Sequences of trpzip and gb1 peptides

trpzip1	SWTWEGNKWTWK	(type II´ turn)	(SEQ ID NO: 1)
trpzip2	SWTWENGKWTWK	(type I´ turn)	(SEQ ID NO: 2)
trpzip3	SWTWEpNKWTWK	(type II´ turn)	(SEQ ID NO: 3)
gb1, 41-56	GEWTYDDATKTFTVTE	(type I turn)	(SEQ ID NO: 4)
trpzip4	GEWTWDDATKTWTWTE	(gb1: Y45W, F52W, V54W)	(SEQ ID NO: 5)
trpzip5	GEWTYDDATKTFTWTE	(gb1: V54W)	(SEQ ID NO: 6)
trpzip6	GEWTWDDATKTWTVTE	(gb1: Y45W, F52W)	(SEQ ID NO: 7)
trpzip7	GEWHWDDATKTWVWTE		(SEQ ID NO: 8)
trpzip8	GEWVWDDATKTWHWTE		(SEQ ID NO: 9)
trpzip9	GEWVWDDATKTWVWTE		(SEQ ID NO: 10)

All peptides were synthesized as C-terminal amides; p ≡ D-proline. Residue numbers for the gb1 peptide correspond to those of the parent 56-residue B1 domain.

Table 2. Thermal unfolding and sedimentation analysis of trpzip peptides

parameter	trpzip1	trpzip2	trpzip3	trpzip4	trpzip5	trpzip6
T_m, K	323.0 ± 0.3	345.0 ± 0.1	351.8 ± 0.2	343.1 ± 0.1	315.8 ± 0.2	317.7 ± 0.48
ΔH_m, cal mol^{-1}	10790 ± 120	16770 ± 60	13020 ± 70	21860 ± 60	13320 ± 140	10290 ± 300
ΔS_m, cal mol^{-1} K^{-1}	33.4	48.6	37.0	63.7	42.2	32.4
ΔC_p, cal mol^{-1} K^{-1}	231 ± 4	281 ± 2	195 ± 2	380 ± 4	325 ± 10	236 ± 17
$\sigma_{obs}/\sigma_{calc}$*	1.02 ± 0.04	1.01 ± 0.04	1.00 ± 0.04	n. d.[†]	n. d.	n. d.

Thermal melts were acquired with 20 µM peptide samples in 20 mM potassium phosphate, pH 7.0. *σ ≡ reduced apparent molecular weight, as determined from sedimentation data fit to a non-ideal single-species model (see Methods). [†]n.d. ≡ not determined; the thermal denaturation curve of trpzip4 was identical at five-fold higher peptide concentration (100 µM vs. 20 µM). Thermal unfolding parameters of $\Delta H = 11600$ cal mol^{-1} and $\Delta S = 39$ cal mol^{-1} K^{-1} have been reported for the gb1 peptide, assuming $\Delta Cp = 0$.

Table 3. NMR structural statistics for trpzip peptides

parameter	trpzip1	trpzip2	trpzip4
R.m.s. deviation from exp'tal distance restraints (Å) (number of restraints)	0.005 ± 0.001 (77)	0.004 ± 0.003 (84)	0.003 ± 0.001 (117)
R.m.s. deviation from exp'tal dihedral restraints (°) (number of restraints)	0.14 ± 0.09 (15)	0.16 ± 0.09 (15)	0.33 ± 0.08 (21)
Maximum distance violation (Å)	0.03 ± 0.00	0.04 ± 0.03	0.03 ± 0.01
Maximum dihedral violation (°)	0.5 ± 0.3	0.6 ± 0.3	1.1 ± 0.3
Ramachandran geometry (% in most favored region)*	71 ± 10	85 ± 10	82 ± 4
Backbone (N,Cα,C) rmsd from mean coordinates (Å) (residues used for rmsd calculation)	0.40 ± 0.07 (2-11)	0.41 ± 0.09 (2-11)	0.29 ± 0.06 (43-54)

Resonance assignments and coupling constants for trpzip1, trpzip2, and trpzip4 are provided in Tables 5-7, respectively. *Ramachandran geometry was evaluated using the program PROCHECK (Laskowski et al. (1993) *J. Appl. Crystallogr.* 26:283-291); remainder of the residues for all structures are in the allowed regions of ϕ,ψ space, with none in the disallowed or generously allowed regions.

Table 4. Thermal unfolding analysis of trpzips 7-9

parameter	trpzip7	trpzip8	trpzip9
T_m, K	353.4 ± 0.1	352.2 ± 0.1	365.1 ± 0.0
ΔH_m, cal mol^{-1}	25030 ± 100	25980 ± 110	26690 ± 80
ΔS_m, cal mol^{-1} K^{-1}	70.8	73.8	73.1
ΔC_p, cal mol^{-1} K^{-1}	418 ± 4	440 ± 4	402 ± 2

Thermal melts were acquired with 20 μM peptide samples in 20 mM potassium phosphate, pH 7.0.

Table 5. Resonance assignments and coupling constants for trpzip1 at 288K, pH 5.5

Res	H^N	H^α	H^β	Other	$^3J_{H\alpha\text{-}H\beta}$	$^3J_{HN\text{-}H\alpha}$
1 Ser	-	3.40	3.69*		NA	
2 Trp	8.81	5.20	<u>3.02</u>, 3.13	$\delta1=7.38$; $\epsilon1=10.28$; $\epsilon3=7.46$; $\zeta2=7.37$; $\zeta3=7.20$; $\eta2=7.26$	11.5, 2.9	NA
3 Thr	9.56	4.85	3.99	$\gamma=1.12$	NA	8.9
4 Trp	8.92	4.61	2.07, <u>2.94</u>	$\delta1=6.96$; $\epsilon1=9.80$; $\epsilon3=5.49$; $\zeta2=7.17$; $\zeta3=6.45$; $\eta2=6.88$	5.5, 10.9[†]	7.9
5 Glu	8.36	4.34	1.75, 1.87	$\gamma=2.01, 2.09$	7.0, 7.1	8.6
6 Gly	8.21	3.48, 3.77				6.7, 6.0
7 Asn	8.14	3.93	2.74, 2.79	$\delta=6.83, 7.50$	NA	8.4
8 Lys	6.53	4.16	1.66, 1.72	$\gamma=1.09, 1.24$; $\delta=1.60*$; $\epsilon=2.95*$	NA	NA
9 Trp	8.55	5.17	<u>2.95</u>, 3.27	$\delta1=7.26$; $\epsilon1=9.93$; $\epsilon3=7.31$; $\zeta2=7.22$; $\zeta3=7.09$; $\eta2=7.17$	10.6, 3.3	7.8
10 Thr	9.77	4.86	4.00	$\gamma=1.15$	NA	NA
11 Trp	9.00	4.26	2.01, <u>2.76</u>	$\delta1=6.80$; $\epsilon1=10.02$; $\epsilon3=5.31$; $\zeta2=7.36$; $\zeta3=6.58$; $\eta2=7.08$	4.6, 12.3	7.4
12 Lys	7.73	4.16	<u>1.37</u>, 1.50	$\gamma=1.14, 1.20$; $\delta=1.49*$; $\epsilon=2.78*$	9.8, 5.0	9.3
13 NH$_2$	6.69, 7.04					

Chemical shifts for the pro R protons of stereospecifically assigned methylene groups are underlined. * indicates degenerate methylene protons. NA indicates that the necessary peak was too overlapped or broad to determine an accurate value of the coupling constant. [†]Note: assuming that the Hα-Hβ coupling constants are a weighted average resulting from the three low-energy $\chi1$ rotamers (-60°, 180°, +60°), then values of 5.5 and 10.9 Hz in conjunction with analysis of local ROEs gives a population distribution for Trp4 $\chi1$ with ratios –60°:180:+60° of approximately 3.5:1:0 [Kessler, H., Griesinger, C., & Wagner, K. (1987) *J. Am. Chem. Soc. 109*, 6927-6933]. An NOE between Trp4$^{H\zeta2}$ and Asn7$^{H\alpha}$ was observed that apparently arises from the small population with Trp4 $\chi1 = 180°$; this NOE is inconsistent with the major –60° $\chi1$ conformation and was removed from the structure calculation to avoid distortion of the turn geometry.

Table 6. Resonance assignments and coupling constants for trpzip2 at 288K, pH 5.5

Res	H^N	H^α	H^β	Other	$^3J_{H\alpha\text{-}H\beta}$	$^3J_{HN\text{-}H\alpha}$
1 Ser	-	3.81	3.54*		NA	
2 Trp	8.91	5.25	3.02, 3.09	$\delta1$=7.39; $\epsilon1$=10.25; $\epsilon3$=7.44; $\zeta2$=7.39; $\zeta3$=7.23; ; $\eta2$=7.30	12.3, 3.1	8.0
3 Thr	9.58	4.91	4.04	γ=1.15	NA	9.4
4 Trp	8.89	4.67	2.01, 2.94	$\delta1$=6.86; $\epsilon1$=9.91; $\epsilon3$=5.66; $\zeta2$=7.22; $\zeta3$=6.55; $\eta2$=6.98	5.9, 10.7[†]	7.9
5 Glu	8.50	4.36	1.78, 1.93	γ=2.05, 2.13	6.2, 8.3	8.2
6 Asn	9.24	4.19	2.65, 2.94	δ=6.97, 7.69	9.1, 6.1	6.4
7 Gly	7.71	3.23, 3.82				6.3, 4.8
8 Lys	6.88	4.26	1.66, 1.72	γ=1.23, 1.28; δ=1.67*; ϵ=3.02*	7.2, 5.1	8.7
9 Trp	8.61	5.17	2.91, 3.28	$\delta1$=7.21; $\epsilon1$=9.81; $\epsilon3$=7.26; $\zeta2$=7.22; $\zeta3$=7.09; $\eta2$=7.19	10.7, 4.0	7.1
10 Thr	9.90	4.90	4.07	γ=1.21	NA	9.8
11 Trp	9.06	4.27	1.95, 2.74	$\delta1$=6.82; $\epsilon1$=10.07; $\epsilon3$=5.25; $\zeta2$=7.41; $\zeta3$=6.60; $\eta2$=7.13	5.0, 12.5	7.9
12 Lys	7.65	4.21	1.40, 1.51	γ=1.19, 1.25; δ=1.55*; ϵ=2.78, 2.85	9.7, 5.6	9.5
13 NH₂	6.70, 7.37					

Chemical shifts for the pro R protons of stereospecifically assigned methylene groups are underlined. * indicates degenerate methylene protons. NA indicates that the necessary peak was too overlapped or broad to determine an accurate value of the coupling constant. [†]Note: assuming that the coupling constants are a weighted average resulting from the three low-energy $\chi1$ rotamers (-60°, 180°, +60°), then values of 5.9 and 10.7 Hz in conjunction with analysis of local ROEs gives a population distribution for Trp4 $\chi1$ with ratios –60°:180:+60° of approximately 3:1:0 [Kessler, H., Griesinger, C., & Wagner, K. (1987) *J. Am. Chem. Soc. 109*, 6927-6933]. An NOE between Trp4$^{H\zeta2}$ and Gly7$^{H\alpha1}$ was observed that apparently arises from the small population with Trp4 $\chi1$ = 180°; this NOE is inconsistent with the major –60° $\chi1$ conformation and was removed from the calculation to avoid distortion of the turn geometry.

Table 7. Resonance assignments and coupling constants for trpzip4 at 288K, pH 6.0

Res	H^N	H^α	H^β	Other	$^3J_{H\alpha\text{-}H\beta}$	$^3J_{HN\text{-}H\alpha}$
41 Gly	-	3.11, 3.55.				
42 Glu	7.38	4.31	1.77, 1.96	γ=1.96*	6.0, 2.7	7.5
43 Trp	8.46	5.36	3.06, 3.40	δ1=7.25; ε1=9.79; ε3=7.60; ζ2=7.20; ζ3=7.23; η2=7.21	10.4, 1.9	8.4
44 Thr	9.92	5.01	4.11	γ=1.21	NA	9.9
45 Trp	9.05	4.24	1.71, 2.63	δ1=6.87; ε1=10.28; ε3=4.97; ζ2=7.21; ζ3=6.27; η2=6.87	5.3, 12.1	7.2
46 Asp	7.85	4.58	2.26, 2.63		4.1, 12.0	9.9
47 Asp	8.61	4.09	2.56, 2.69		8.5, 6.9	5.6
48 Ala	8.25	4.17	1.49			5.9
49 Thr	7.11	4.17	4.20	γ=1.09	NA	9.2
50 Lys	7.71	2.99	1.86, 2.17	γ=1.28, 1.38; δ=1.75*; ε=3.10*	3.8, 12.1	7.5
51 Thr	6.40	4.43	3.93	γ=1.10	5.3	8.8
52 Trp	8.29	5.30	2.99, 3.38	δ1=7.17; ε1=9.74; ε3=7.58; ζ2=7.10; ζ3=7.21; η2=7.18	10.8, 1.8	9.7
53 Thr	9.80	5.00	4.07	γ=1.20	NA	9.8
54 Trp	9.03	4.50	1.75, 2.71	δ1=6.73; ε1=10.03; ε3=5.21; ζ2=7.25; ζ3=6.40; η2=6.95	4.7, 12.2	7.7
55 Thr	8.16	4.26	3.89	γ=1.05	5.3	9.5
56 Glu	8.39	3.88	1.87, 2.02	γ=2.30*	8.9, 6.2	6.6
57 NH₂	7.12, 7.56					

Chemical shifts for the pro R protons of stereospecifically assigned methylene groups are underlined. * indicates degenerate methylene protons. NA indicates that the necessary peak was too overlapped or broad to determine an accurate value of the coupling constant.

While the invention has necessarily been described in conjunction with preferred embodiments, one of ordinary skill, after reading the foregoing specification, will be able to effect various changes, substitutions of equivalents, and alterations to the subject matter set forth herein, without departing from the spirit and scope thereof. Hence, the invention can be practiced in ways other than those specifically described herein.